Estatística não paramétrica

Professor: Dr. Pedro M. Almeida-Junior

18 de março de 2023

Departamento de Estatística (UEPB)

TESTES PARA O CASO DE UMA AMOSTRA (Testes de aderência)

Testes de uma amostra

- O teste de uma amostra é, em geral, uma prova de aderência. No caso típico, extraímos uma amostra aleatória de alguma população e então testamos a hipótese de que a amostra tenha sido extraída de uma população com distribuição específica.

- **Os testes de uma amostra podem responder questões como estas:**
  - ✓ Há diferença significante entre a posição (tendência central) da amostra e da população?
  - ✓ Há diferença significante entre as frequências observadas e as frequências que esperaríamos com base na mesma teoria?
  - ✓ Há diferença significante entre proporções observadas e esperadas em uma sequência de observações dicotômicas?
  - ✓ É razoável acreditar que a amostra tenha sido extraída de uma população com uma distribuição específica (por exemplo, normal ou uniforme)?

Testes de aderência

Estes testes são úteis para verificar se determinada amostra pode provir de uma população especificada. São usualmente conhecidos como testes de aderência ou bondade do ajuste. Neste caso, retira-se uma amostra aleatória e compara-se a distribuição amostral com a distribuição de interesse.

Principais Testes:

- Teste Binomial
- Teste de Qui-Quadrado (χ^2)
- Teste de Kolmogorov-Srnirnov
- Teste de Lilliefors
- Teste de Aleatorização

Teste Binomial

- É aplicado em amostras provenientes de populações que constituem-se de apenas 2 categorias (variáveis dicotômicas).
- Exemplo: (negativo, positivo) ou (defeituoso, não defeituoso)
- É útil para verificarmos se a proporção de sucesso p observada na amostra pode pertencer a uma população com um determinado valor de p.

Pressupostos

- Cada observação é classificada como sucesso ou fracasso;
- A probabilidade p de sucesso não se altera com a repetição do experimento;
- As n tentativas são independentes.

Este método irá calcular a probabilidade de obter valores mais extremos do que os observados.

- Inicia-se com a formulação das hipóteses:

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 \; ou \; H_1 : p < p_0 \text{ (unilateral) ou } H_1 : p \neq p_0 \text{(bilateral)} \end{cases}$$

- Determinar n
- Determinar as Frequências observadas
- De acordo com as hipóteses, usando a distribuição Binomial, calculamos a probabilidade de ocorrência de valores tão ou mais extremos do que o observado (número de sucessos) $-p$-valor:

$$P(X \leq x) = \sum_{i=0}^{x} \binom{n}{i} p^i q^{n-i} \text{ ou } P(X \geq x) = \sum_{i=0}^{x} \binom{n}{i} p^i q^{n-i} \tag{1}$$

Exemplo 1: Uma ilustração simples esclarecerá a Equação 1. Suponha que um dado honesto é lançado cinco vezes. Qual é a probabilidade de que exatamente dois dos lançamentos mostrem "seis"? Nesse caso Y é a variável aleatória (o resultado de cinco lançamentos do dado), $N =$ o número de lançamentos $(5), k =$ o número observado de seis $(2), p =$ a proporção esperada de seis $\left(\frac{1}{6}\right)$.
Qual a probabilidade de que exatamente dois dos cinco lançamentos venham a mostrar a face seis ?

Usando a Equação 1, então

$$P[Y = k] = \binom{N}{k} p^k (1 - p)^{N-k}$$

Usando a Equação 1, então

$$P[Y = k] = \binom{N}{k} p^k (1-p)^{N-k}$$

$$P[Y = 2] = \tfrac{5!}{2!3!} \left(\tfrac{1}{6}\right)^2 \left(\tfrac{5}{6}\right)^3 = 0,16$$

Suponha que queiramos saber a probabilidade de obter dois ou menos "seis"quando um dado honesto é lançado cinco vezes. Agora a probabilidade de obter, no máximo, dois "seis"é denotada por $P[Y \leq 2]$. Da Equação 1 a probabilidade de obter 0 "seis"é $P[Y = 0]$, a probabilidade de obter um "seis"é $P[Y = 1]$, etc. Temos então

$$P[Y \leq 2] = P[Y = 0] + P[Y = 1] + P[Y = 2]$$

Isto é, a probabilidade de se obter no máximo dois "seis"é a soma de três probabilidades. Então,

$$P[Y = 0] =$$

$$P[Y = 1] =$$

$$P[Y = 2] =$$

Suponha que queiramos saber a probabilidade de obter dois ou menos "seis"quando um dado honesto é lançado cinco vezes. Agora a probabilidade de obter, no máximo, dois "seis"é denotada por $P[Y \leq 2]$. Da Equação 1 a probabilidade de obter 0 "seis"é $P[Y = 0]$, a probabilidade de obter um "seis"é $P[Y = 1]$, etc. Temos então

$$P[Y \leq 2] = P[Y = 0] + P[Y = 1] + P[Y = 2]$$

Isto é, a probabilidade de se obter no máximo dois "seis"é a soma de três probabilidades. Então,

$$P[Y = 0] = \tfrac{5!}{0!5!} \left(\tfrac{1}{6}\right)^0 \left(\tfrac{5}{6}\right)^5$$

$$P[Y = 1] =$$

$$P[Y = 2] =$$

Suponha que queiramos saber a probabilidade de obter dois ou menos "seis"quando um dado honesto é lançado cinco vezes. Agora a probabilidade de obter, no máximo, dois "seis"é denotada por $P[Y \leq 2]$. Da Equação 1 a probabilidade de obter 0 "seis"é $P[Y = 0]$, a probabilidade de obter um "seis"é $P[Y = 1]$, etc. Temos então

$$P[Y \leq 2] = P[Y = 0] + P[Y = 1] + P[Y = 2]$$

Isto é, a probabilidade de se obter no máximo dois "seis"é a soma de três probabilidades. Então,

$$P[Y = 0] = \tfrac{5!}{0!5!} \left(\tfrac{1}{6}\right)^0 \left(\tfrac{5}{6}\right)^5 = 0,40$$

$$P[Y = 1] =$$

$$P[Y = 2] =$$

Suponha que queiramos saber a probabilidade de obter dois ou menos "seis"quando um dado honesto é lançado cinco vezes. Agora a probabilidade de obter, no máximo, dois "seis"é denotada por $P[Y \leq 2]$. Da Equação 1 a probabilidade de obter 0 "seis"é $P[Y = 0]$, a probabilidade de obter um "seis"é $P[Y = 1]$, etc. Temos então

$$P[Y \leq 2] = P[Y = 0] + P[Y = 1] + P[Y = 2]$$

Isto é, a probabilidade de se obter no máximo dois "seis"é a soma de três probabilidades. Então,

$$P[Y = 0] = \tfrac{5!}{0!5!} \left(\tfrac{1}{6}\right)^0 \left(\tfrac{5}{6}\right)^5 = 0,40$$

$$P[Y = 1] = \tfrac{5!}{1!4!} \left(\tfrac{1}{6}\right)^1 \left(\tfrac{5}{6}\right)^4$$

$$P[Y = 2] =$$

Suponha que queiramos saber a probabilidade de obter dois ou menos "seis"quando um dado honesto é lançado cinco vezes. Agora a probabilidade de obter, no máximo, dois "seis"é denotada por $P[Y \leq 2]$. Da Equação 1 a probabilidade de obter 0 "seis"é $P[Y = 0]$, a probabilidade de obter um "seis"é $P[Y = 1]$, etc. Temos então

$$P[Y \leq 2] = P[Y = 0] + P[Y = 1] + P[Y = 2]$$

Isto é, a probabilidade de se obter no máximo dois "seis"é a soma de três probabilidades. Então,

$$P[Y = 0] = \tfrac{5!}{0!5!} \left(\tfrac{1}{6}\right)^0 \left(\tfrac{5}{6}\right)^5 = 0,40$$

$$P[Y = 1] = \tfrac{5!}{1!4!} \left(\tfrac{1}{6}\right)^1 \left(\tfrac{5}{6}\right)^4 = 0,40$$

$$P[Y = 2] =$$

Suponha que queiramos saber a probabilidade de obter dois ou menos "seis"quando um dado honesto é lançado cinco vezes. Agora a probabilidade de obter, no máximo, dois "seis"é denotada por $P[Y \leq 2]$. Da Equação 1 a probabilidade de obter 0 "seis"é $P[Y = 0]$, a probabilidade de obter um "seis"é $P[Y = 1]$, etc. Temos então

$$P[Y \leq 2] = P[Y = 0] + P[Y = 1] + P[Y = 2]$$

Isto é, a probabilidade de se obter no máximo dois "seis"é a soma de três probabilidades. Então,

$$P[Y = 0] = \tfrac{5!}{0!5!} \left(\tfrac{1}{6}\right)^0 \left(\tfrac{5}{6}\right)^5 = 0,40$$

$$P[Y = 1] = \tfrac{5!}{1!4!} \left(\tfrac{1}{6}\right)^1 \left(\tfrac{5}{6}\right)^4 = 0,40$$

$$P[Y = 2] = \tfrac{5!}{2!3!} \left(\tfrac{1}{6}\right)^2 \left(\tfrac{5}{6}\right)^3$$

Suponha que queiramos saber a probabilidade de obter dois ou menos "seis"quando um dado honesto é lançado cinco vezes. Agora a probabilidade de obter, no máximo, dois "seis"é denotada por $P[Y \leq 2]$. Da Equação 1 a probabilidade de obter 0 "seis"é $P[Y = 0]$, a probabilidade de obter um "seis"é $P[Y = 1]$, etc. Temos então

$$P[Y \leq 2] = P[Y = 0] + P[Y = 1] + P[Y = 2]$$

Isto é, a probabilidade de se obter no máximo dois "seis"é a soma de três probabilidades. Então,

$$P[Y = 0] = \tfrac{5!}{0!5!} \left(\tfrac{1}{6}\right)^0 \left(\tfrac{5}{6}\right)^5 = 0,40$$

$$P[Y = 1] = \tfrac{5!}{1!4!} \left(\tfrac{1}{6}\right)^1 \left(\tfrac{5}{6}\right)^4 = 0,40$$

$$P[Y = 2] = \tfrac{5!}{2!3!} \left(\tfrac{1}{6}\right)^2 \left(\tfrac{5}{6}\right)^3 = 0,16$$

Exemplo 2: Uma empresa lançou recentemente no mercado uma dieta de emagrecimento. E pretende-se averiguar se a proporção de seguidores com idades superiores a 60 anos é superior a 0.3. Para tal procedeu na escolha aleatória de 16 indivíduos entre os adeptos da nova dieta, e apurou que apenas 5 tinham idade superior a 60 anos.

1. Hipóteses:

1. Hipóteses:$H_0 : p \leq 0,30 \times H_1 : p > 0,30$

1. Hipóteses:$H_0 : p \leq 0,30 \times H_1 : p > 0,30$
2. $x = 5$(quantidade de sucessos)

1. Hipóteses:$H_0 : p \leq 0,30 \times H_1 : p > 0,30$
2. $x = 5$(quantidade de sucessos)
3. $n = 16$ (amostra)

1. Hipóteses: $H_0 : p \leq 0,30 \times H_1 : p > 0,30$
2. $x = 5$(quantidade de sucessos)
3. $n = 16$ (amostra)
4.

$$\begin{aligned} P(X \geq 5) &= \sum_{i=5}^{16} \left(\begin{array}{c} 16 \\ i \end{array} \right) (0,3)^i (0,7)^{16-i} \\ &= \end{aligned} \tag{2}$$

1. Hipóteses: $H_0 : p \leq 0,30 \times H_1 : p > 0,30$
2. $x = 5$ (quantidade de sucessos)
3. $n = 16$ (amostra)
4.

$$
\begin{aligned}
P(X \geq 5) &= \sum_{i=5}^{16} \begin{pmatrix} 16 \\ i \end{pmatrix} (0,3)^i (0,7)^{16-i} \\
&= 1 - P(X \leq 4)
\end{aligned}
\tag{2}
$$

1. Hipóteses:$H_0 : p \leq 0,30 \times H_1 : p > 0,30$
2. $x = 5$(quantidade de sucessos)
3. $n = 16$ (amostra)
4.

$$\begin{aligned} P(X \geq 5) &= \sum_{i=5}^{16} \left(\begin{array}{c} 16 \\ i \end{array} \right) (0,3)^i (0,7)^{16-i} \\ &= 1 - P(X \leq 4) \\ &= 1 - \sum_{i=0}^{4} \left(\begin{array}{c} 16 \\ i \end{array} \right) (0,3)^i (0,7)^{16-i} = \end{aligned} \tag{2}$$

1. Hipóteses:$H_0 : p \leq 0,30 \times H_1 : p > 0,30$
2. $x = 5$(quantidade de sucessos)
3. $n = 16$ (amostra)
4.

$$\begin{aligned} P(X \geq 5) &= \sum_{i=5}^{16} \left(\begin{array}{c} 16 \\ i \end{array} \right) (0,3)^i (0,7)^{16-i} \\ &= 1 - P(X \leq 4) \\ &= 1 - \sum_{i=0}^{4} \left(\begin{array}{c} 16 \\ i \end{array} \right) (0,3)^i (0,7)^{16-i} = 0,55 \end{aligned} \tag{2}$$

1. Hipóteses:$H_0 : p \leq 0,30 \times H_1 : p > 0,30$
2. $x = 5$(quantidade de sucessos)
3. $n = 16$ (amostra)
4.

$$\begin{aligned} P(X \geq 5) &= \sum_{i=5}^{16} \binom{16}{i} (0,3)^i (0,7)^{16-i} \\ &= 1 - P(X \leq 4) \\ &= 1 - \sum_{i=0}^{4} \binom{16}{i} (0,3)^i (0,7)^{16-i} = 0,55 \end{aligned} \tag{2}$$

1. Hipóteses:$H_0 : p \leq 0,30 \times H_1 : p > 0,30$
2. $x = 5$(quantidade de sucessos)
3. $n = 16$ (amostra)
4.

$$
\begin{aligned}
P(X \geq 5) &= \sum_{i=5}^{16} \left(\begin{array}{c} 16 \\ i \end{array} \right) (0,3)^i (0,7)^{16-i} \\
&= 1 - P(X \leq 4) \\
&= 1 - \sum_{i=0}^{4} \left(\begin{array}{c} 16 \\ i \end{array} \right) (0,3)^i (0,7)^{16-i} = 0,55
\end{aligned} \tag{2}
$$

Conclusão: Com um nível de significância ($\alpha = 0,05$) não rejeitamos a hipótese nula pois o p-valor é aproximadamente igual a 0,55. Portanto, não podemos afirmar que a proporção de seguidores com idades superiores a 60 anos é superior a 0,30.

Discussão do Método

É a técnica mais poderosa aplicável a dados medidos em escala nominal e ainda por cima dicotômicas. Quando a variável aleatória em estudo é contínua e procede-se uma dicotomização haverá certamente perda de eficiência.